

Weekly Report

Done

1. 经与巫老师讨论, BeXplorer 最好就投给 TIST. 仅对一些比较简单的建议作出改动便投出:
 - a. 强行添加游戏相关文献 |之前删除的一部分也补回来
 - b. 调整三个概念定义的顺序
 - c. 解释视图里面一些细节 (reviewer 1)
 - d. 补回之前删掉的社会学上的一些 generalizability 以及解释, 如果篇幅可以的话还可以补回 case
2. 与老师同学们一些简单的讨论, 主要概述如下:
 - a. 核心思想是找到一个新的小问题\小切入点, 并结合一些先进的技术去解决.
 - b. 最好是已有数据的或者我们能找到的或者能通过简单实验大量得到的
 - c. 细节方面
 - i. Graph 上. Matrix 表示方法与 node-link 的选择, 能否从图像空间去解决一些传统方法速度较慢的问题 (如 motif search, 不同的列排布顺序展现不同的 motif pattern; 或是其他一些 frequent subgraph mining 这样的. 注意到, 这两种 pattern 类型并不一致, 后者更有一种 meta 的感觉);
 1. Nodelink matrix 与数据本身直接三个空间能否有一种潜在关联?
 - ii. 对于海东师兄的蓝噪声采样, 我们能否用于线(如轨迹\PCP)或者网络可视化上(周志光老师的采样仍然是在点上做文章反推到线段上的).
 1. 由于这边仍然需要可视化结果上的这里自然能想到的是 autoencoder 以及各类亲戚, 也就是先压缩再还原的方法, 本身来说也是一种压缩, 并把最有用最本质最核心的东西保留出来.
 2. 蓝噪声相关的知识需要补充一下, 最终是一种优化问题.
 - iii. 上述问题都很巧的提到了采样, 大致总结一下
 1. 简单的就是蒙特卡洛及其亲戚
 2. 蓝噪声采样以及其亲戚
 3. Autoencoder 及亲戚, 以及借用其思想的变分 autoencoder
3. 此外本周还在浮光掠影的探索 GAN 及其应用(通过 Ian Goodfellow 一些视频), graph convolution network (GCN)等以及一些数学知识的回顾. 对于这些高级的神经网络方法, 虽然仍然有点玄学在里面, 但还是有一些看起来有点费劲的概率论知识在里面, 比一般的 RNN CNN 等方法会难一点. 并往往和已有一些方法沾上边所以顺带看了那些方法. 在跑实例方面, 还没动手但是花了一些时间装好了框架. 并记录了一些任务准备让小顾练练手.
 - a. 为什么用 GCN? 传统的 CNN 不能在这种 Non Euclidean structure 数据上做东西, 就不能像卷积在图片或者在时间序列一样提取有效特征. 所以要有 GCN, 理论上, 他能干的事情很多, 我们联系一下谱聚类就知道, 很多东西只要能搞出一个关系就能做一个网络然后运用相关的方法, 比如图像似乎也有这种做法, LIDAR 这种自动驾驶常见的 point cloud 也可以?
 - b. 图卷积(基础 GCN 的方法, 输入是节点特征矩阵 + 网络结构的邻接矩阵, 输出是节点特征矩阵, 也可以当做 embedding; 但是输出上通过卷积得到一个 graph-level 的特征), 也可以做分类, 可以做推荐系统, 做影像. 并且这种网络还不是简单的节点\边都固定是一类物体\关系的网络, 往往是那种有各种复杂节点属性和 relationship 的网

络. 今年 VIS18 也看到了有些文章在做相关这种形式的 multivariate graph 可视化的工作. 我个人认为此处也是需要考量的一个点.

To Do

- 1. 从课件或者教材或者讨论中发现问题. 尤其是上面说的一些和采样相关的办法; 继续浮光掠影的补充知识, 但相信现在即便浮光掠影, 看的机器学习内容也一直在收敛中.
 - a. Motif search 相关知识了解下, 现有文献
 - b. 一些采样的方法
 - c. 上面提到一些深度模型
 - d. 将可视化问题适配给上面一些方法. 聚焦在简单的点\线\网三处挖掘问题.
- 2. 开始改改 BeXplorer

论文阅读

理解 Mahalanobis distance 以及其与 PCA 关系以及为什么要用这种距离; 了解 VAE 一些深层数学背景, GAN 以及早期的一些应用, GCN 等等.

Task taxonomy for graph visualization 可视化中的图相关的一些 task. 内容比较基础, 略有一些启发. 有一些相关工作可以作为上面提到的矩阵排序的初步探索. 此外, 文中一些设计所有可视化来分析任务的 task taxonomy 也是一个可以用于参考的指标.

The Perception of Graph Properties in Graph Layouts **EUROGRAPHICS18** 对于三种简单的 layout 以及一些常见的网络指标进行 user study, 看是否满足韦伯定律. 作为一个试水工作, 感觉这块很有搞头, 不仅指导了可视化设计, 留下许多可以 user study 的坑, 甚至也可以套上马哥那些用 NN 模拟感知的方法来研究相关问题.

Juniper: A Tree+Table Approach to Multivariate Graph Visualization **VIS18** 对于 multivariate graph 做可视化, 利用的是 spanning tree 技术.

Efficient Reflectance Capture Using an Autoencoder **SIGGRAPH18** 图形学中的 BRDF 表示的是双向反射分布函数 (Bidirectional Reflectance Distribution Function), 它描述了光线如何在物体表面进行反射, 可以用来描述材质属性. 本文利用较少的样本(相比其他同类方法), 利用比较简单的 autoencoder(encoder 是天然的, 硬件相关的, decoder 是自己设计的)来还原 BRDF 的数学形式.

工作时间

平时 10, 周末一共 6 小时. 总共 56.

个人规划

性质	Ddl	进展	目标
----	-----	----	----

中长期	2019 年中	MCMC, 矩阵分解与特征值到底为什么有用, 傅里叶变换	统计 优化 算法
短期	18 年底	浮光掠影 GAN, GCN, VAE 背后的理论假设. 一些方法之间彼此的关联	各种学习对可视化支持
短期	18 年底	设定目标	BeXplorer 投 TIST
长期			博士毕业论文中的研究方向